



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Misperceived social norms: women working outside the home in Saudi Arabia

Bursztyn, Leonardo ; González, Alessandra L ; Yanagizawa-Drott, David

Abstract: We show that the vast majority of young married men in Saudi Arabia privately support women working outside the home (WWOH) and substantially underestimate support by other similar men. Correcting these beliefs increases men's (costly) willingness to help their wives search for jobs. Months later, wives of men whose beliefs were corrected are more likely to have applied and interviewed for a job outside the home. In a recruitment experiment with a local company, randomly informing women about actual support for WWOH leads them to switch from an at-home temporary enumerator job to a higher-paying, outside-the-home version of the job.

DOI: <https://doi.org/10.1257/aer.20180975>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-195317>

Journal Article

Published Version

Originally published at:

Bursztyn, Leonardo; González, Alessandra L; Yanagizawa-Drott, David (2020). Misperceived social norms: women working outside the home in Saudi Arabia. *American Economic Review*, 110(10):2997-3029.

DOI: <https://doi.org/10.1257/aer.20180975>

Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia[†]

By LEONARDO BURSZTYN, ALESSANDRA L. GONZÁLEZ,
AND DAVID YANAGIZAWA-DROTT*

We show that the vast majority of young married men in Saudi Arabia privately support women working outside the home (WWOH) and substantially underestimate support by other similar men. Correcting these beliefs increases men's (costly) willingness to help their wives search for jobs. Months later, wives of men whose beliefs were corrected are more likely to have applied and interviewed for a job outside the home. In a recruitment experiment with a local company, randomly informing women about actual support for WWOH leads them to switch from an at-home temporary enumerator job to a higher-paying, outside-the-home version of the job. (JEL D83, J16, J22, O15, Z13)

Saudi Arabia is a salient example of a country with a low rate of female labor force participation. Less than 15 percent of the Saudi female population aged 15 and above were employed in 2017 (a labor force participation rate of around 18 percent).¹ The share of women working *outside the home*, an action that others can easily observe, is substantially lower. In a national survey we conducted with young married Saudi males in early 2018, only 4 percent had their wife currently working outside the home. Moreover, through the custom of male guardianship, husbands commonly have the final word on their wives' labor supply decisions in Saudi Arabia.

In this paper, we consider opinions on whether women *should* be allowed to work outside the home, and pursue the idea that Saudi society is experiencing “pluralistic ignorance” (Katz, Allport, and Jenness 1931). This refers to a situation where most people privately hold an opinion, but they incorrectly believe that most other people

*Bursztyn: University of Chicago and NBER (email: bursztyn@uchicago.edu); González: University of Chicago (email: alg2@uchicago.edu); Yanagizawa-Drott: University of Zürich (email: david.yanagizawa-drott@econ.uzh.ch). Stefano DellaVigna was the coeditor for this article. We would like to thank four anonymous referees, Abhijit Banerjee, Marianne Bertrand, Ruben Durante, Claudia Goldin, Georgy Egorov, Eliana La Ferrara, Rohini Pande, Ricardo Perez-Truglia, Hans-Joachim Voth, and numerous seminar participants for comments and suggestions, and Hussein Elkheshen, Aakaash Rao, Erik Torstensson, Rebecca Wu, and especially Raymond Han for outstanding research assistance. Our study was approved by the University of Chicago Institutional Review Board. This research was sponsored in part by a grant from the Harvard Kennedy School Evidence for Policy Design program and the Human Resources Development Fund of the Kingdom of Saudi Arabia. The experiments and follow-up survey reported in this study can be found in the AEA RCT Registry (0002447, 0002633, and 0005321).

[†]Go to <https://doi.org/10.1257/aer.20180975> to visit the article page for additional materials and author disclosure statements.

¹See General Authority for Statistics, Kingdom of Saudi Arabia (2017).

hold the contrary opinion, and therefore end up acting against their own view. When individuals believe a behavior or attitude is stigmatized, they might be reluctant to reveal their private views to others for fear of social sanction.² If most individuals act this way, they might all end up believing their private views are only shared by a small minority at most.³ We, therefore, ask: do Saudi men have correct perceptions of the opinions held by other men regarding women working outside the home (WWOH)? If the social norm is misperceived, then does correcting beliefs lead more women to work outside the home?

We combine experiments and surveys to first provide incentivized evidence that the majority of married men in Saudi Arabia in fact support WWOH, while they substantially underestimate the level of support for WWOH by other men—even those from their same social setting, such as neighbors. We then show that randomly correcting these beliefs about others increases married men's willingness to let their wives search for jobs (as measured by their costly sign up for a mobile job matching service for their wives). We also find that this decision maps onto real outcomes. Three to five months after the main intervention, the wives of men in our original sample (whose beliefs about acceptability of WWOH were corrected) are more likely to have applied and interviewed for a job outside the home. Finally, using a natural recruitment experiment with a local company, we find that misperception correction leads to a higher share of women taking a higher-paying, outside-the-home temporary survey enumerator job over a similar job to be performed from home.

More specifically, we first report on an experiment with a sample of 500 Saudi married men aged 18–35, recruited from different neighborhoods in Riyadh. These men each attended a 30-participant session, composed of individuals from the same geographical area, thus sharing a common social network.⁴ In an anonymous online survey, 87 percent of the experimental participants agreed with the statement, “In my opinion, women should be allowed to work outside the home.” When incentivized to guess how other session participants responded to the same question, about three-quarters of the experimental subjects *underestimate* the true number. We interpret this as evidence of misperception of social norms, even among people from the same neighborhood who know each other.

Next, we evaluate whether correcting these misperceptions matters for a revealed preference decision, associated with household labor supply. One-half of the participants were randomly given feedback on the true number of agreements with the statement in their session. At the end of the experiment, subjects were asked

²Following Bénabou and Tirole (2011), we think of social norms as the set of “social sanctions or rewards” that incentivize a certain behavior. We examine *injunctive*, but not *descriptive* norms.

³Historic examples of pluralistic ignorance include the late Soviet regime (Kuran 1991), where many individuals opposed the regime but believed others supported it. In 1968, most white Americans substantially overestimated the support for racial segregation among other whites (O’Gorman 1975). Work in psychology has also documented pluralistic ignorance regarding alcohol use on college campuses (Prentice and Miller 1993). A related concept is “preference falsification” (Kuran 1995): people’s stated, public preferences are influenced by social acceptability, and might be different from their true, private preferences. More closely related to our study, González (2013) documents that the majority of male Kuwaiti college students in her sample believed women should work outside the home, while they thought that the majority of their religious community would not approve of it. For a recent overview in social psychology, see Tankard and Paluck (2016). For the related concept of “third-order inference” in sociology, see Correll et al. (2017). Recent work by Bursztyn, Egorov, and Fiorin (forthcoming) has analyzed how factors correcting beliefs about what others think might generate fast changes in the perceived acceptability of certain behaviors, and also in actual behavior.

⁴The average participant reported knowing 15 of the 29 other session participants.

to make an incentivized choice between receiving an additional bonus payment (an online gift card) and signing their wives up for a job matching mobile application, specializing in the Saudi female labor market. In a control group without belief corrections, 23 percent of participants chose the job matching service. In the treatment group with feedback on the opinions of other participants, the share went up significantly, by 9 percentage points (a 36 percent increase). The increase is driven by those who underestimated the true share of WWOH participant supporters in their session: sign up rates go up by 57 percent (from a baseline rate of 21 percent) when this group is provided with information, while information doesn't change sign up rates for those who did not underestimate support by others (that group also has a higher baseline sign up rate of 31 percent).⁵

One might be worried that the sign up outcome is not strongly indicative of actual labor supply decisions, or that the immediate decision does not imply a more permanent change in perceived social norms and behavior. To deal with these concerns, three to five months after the original intervention, participants were recontacted by phone and a series of additional outcome questions were collected.⁶ We document a longer-term impact on self-reported labor supply outcomes. Wives of treated participants were significantly more likely to have applied for a job outside the home (up by 10 percentage points from a baseline level of 6 percent) and to have interviewed for a job outside the home (up by 5 percentage points from a baseline level of 1 percent). We are not powered to detect a significant change in the likelihood of the wife being employed outside the home, though we directionally observe an increase. We also document that the change in perceived social norms is persistent. Treated participants believe a significantly higher share of their neighbors in general support women working outside the home. Finally, we observe that the persistent change in perceived social norms might spill over to other behaviors. Treated participants are significantly more likely to report that they would sign up their wives for driving lessons. These findings are robust to adjustments for multiple hypothesis testing.⁷

⁵This first outcome is a decision made by husbands/guardians, and not by wives themselves. We think husbands' decisions are crucial, since we are examining men's potential reluctance to let their wives join the labor force due to perceived social norms as an obstacle for WWOH. Moreover, due to the custom of male guardianship, husbands typically have the final word on their wives' labor supply decisions. Until 2011, guardians' permission was legally required but since then their permission is asked by many, but not all, hiring firms. In our separate recruitment experiment below, we provide evidence from a firm which does not require written approval. We analyze the effects of giving information directly to women. Also note, that because a participant's wife's eventual employment status is observable, the observability of the sign up choice itself does not matter independently.

⁶To preserve anonymity in the original experiment, while still being able to contact participants in the future, phone numbers were first collected at the session level without matching them to specific respondents, before the experiment started. In the follow-up phone survey, participants were asked for the last three digits of their phone numbers. We were able to match 95 percent of the phone numbers to the combination of last three digits and session number.

⁷Since the vast majority of men in our sample privately support WWOH, we believe updates in perceived social acceptability are the main mechanism driving our findings (in online Appendix A, we present a simple model of labor supply and stigma based on this mechanism). However, it is possible that the information provided leads some participants from having a privately negative opinion about WWOH to a positive one. We did not collect updated opinions after the information was provided to verify this possibility. Still, we can check the treatment effects for those originally opposed to WWOH. We find a large point estimate (10.9 percentage point increase from a baseline sign up rate of 10.7 percent in the control group), but this estimate is not significant, perhaps due to the small sample size for that group ($N = 65$). This might be interpreted as suggestive evidence of a persuasive effect of the information treatment in that subsample. However, it is also consistent with these participants not changing their opinions, but changing their behavior because they care strongly about their social image and their perceptions of norms have been substantially updated. Indeed, the average wedge among those originally negative about WWOH is substantially and significantly larger than among those originally positive: -11 percentage points versus

We also conduct a similar-looking, anonymous online survey with a larger, national sample of about 1,500 married Saudi men, aged 18–35. The goal of this additional survey is twofold. First, we assess the external validity of the finding that most Saudi men privately support WWOH while failing to understand that others do as well. In this broader, more representative sample, 82 percent of men agree with the same statement on WWOH used in the main experiment. When incentivized to guess the responses of other survey respondents, 92 percent of them underestimate the true share. These are stronger misperceptions, perhaps because they are no longer asked about their own neighbors' opinions.

We examine whether social desirability bias/experimenter demand effects could be a driver of the misperception finding in the main experiment. Although the experiment was anonymous, it is possible that some participants may have felt like they had to answer the question about their own views in a certain way. One-half of the online survey participants were assigned an elicitation procedure that provided a “cover” for their opinion on WWOH. In particular, we implemented a “list experiment” (also called the “unmatched count” and the “item count technique,” originally formalized by Raghavarao and Federer 1979).⁸ Using a method that provides respondents with a higher degree of plausible deniability, we find a very similar level of agreement with the statement regarding whether women should be allowed to work outside the home: 80 percent. Finally, we examine whether individuals may incorrectly expect others to strategically respond to the WWOH agreement question, which would distort guesses about others since the question asked about how others *answered* the question. We find that beliefs about other participants' *true opinions* were extremely similar to the guesses about others' *answers*. In sum, we find no evidence that survey response biases drive our results.

As an additional check on the external validity of the fact we document, we show that the share of Saudi men supportive of WWOH is also very similar when using the nationally representative sample from the wave of the *Arab Barometer* containing that question for Saudi Arabia (2010–2011). Out of approximately 700 male respondents, 77 percent are in favor of WWOH and among male respondents aged 18–35 (the age bracket in our study), the share is 79 percent. The *Arab Barometer* survey also allows us to establish that older men are also supportive of WWOH: among those over 35, the share agreeing with that statement is 74 percent. Moreover, the numbers from the *Arab Barometer* in 2010–2011 suggest that the misperception in

–6.7 percentage points. ($p = 0.000$). Furthermore, the point estimate of the treatment effect is unchanged when we restrict the analysis to those who originally reported positive own views about WWOH. Finally, the treatment effects are stronger for those who experience a larger update in beliefs about the opinions of others. As we discuss below, results from an additional survey we conducted provide evidence consistent with our mechanism: most men report very rarely discussing the topic of whether women should be allowed to work outside the home with their male friends and relatives, and those who rarely discuss the subject have much stronger misperceptions than those who commonly discuss it.

⁸The list experiment works as follows: first, respondents are (randomly) assigned either into a control group or to one or to a treatment group. Subjects in all conditions are asked to indicate the *number* of policy positions they support from a list of positions on several issues. Support for any particular policy position is never indicated, only the total number of positions articulated on the list that a subject supports. In the control condition, the list includes a set of contentious, but *not* stigmatized, policy positions. In the treatment condition, the list includes the contentious policy positions from the control list, but also adds the policy position of interest (support for WWOH), which is potentially stigmatized. The degree of support for the stigmatized position *at the sample level* is determined by comparing the average number of issues supported in the treatment and control conditions. For another recent application of list experiments, see Enikolopov et al. (2020).

social norms might have not have been a short-lived phenomenon, at least in the sense that support for WWOH has been high and remained relatively constant until today.

We also provide a discussion on lack of communication as a potential driver for these misperceptions. In an additional national survey we conduct, a large fraction of respondents report to either rarely or very rarely discuss with their male friends and relatives the topic of whether women should be allowed to work outside the home. Moreover, frequency of discussion is a very strong predictor of the size of misperceptions. On average, there are very small misperceptions among those frequently discussing the topic, and very large misperceptions among those rarely talking about it.

Our results carry some natural policy implications. They suggest that there are too few women working outside of the home because the labor market is in an equilibrium where social norms are misperceived—pluralistic ignorance—and that a low-cost intervention correcting perceptions will lead to more women choosing jobs outside the home. To test this directly, we implemented an additional experiment in a *natural* labor market setting, in partnership with a large Saudi survey company. The company embedded an experiment in their efforts to recruit enumerators for a one-day surveyor job. Using a sample of around 300 females that had previously indicated interest in working as a surveyor making calls *from home*, the company offered them the choice between that version of the job and a different version of the same job, interviewing respondents *outside the home*, face-to-face, in malls, for a 20 percent higher wage (with transportation costs additionally covered by the company). Before making the decision of what version of the job to take, the survey firm informed a random set of these job-seekers of the findings from our national survey: “In a recent survey of a national sample of about 1,500 married Saudi men aged 18–35, 82 percent agreed with the statement, “In my opinion, women should be allowed to work outside the home.” This means that the vast majority of young married Saudi men support women working outside of the home.” We are thus able to embed a natural field experiment into an actual hiring process. Moreover, we hold fixed many characteristics of the job (the employer, the nature of the task broadly, including the survey they have to conduct), while only varying whether the job is to be performed at home or outside the home.

The evidence shows that a simple intervention in a natural setting like this can affect labor supply decisions. Information significantly increase the likelihood that job-seekers choose the outside the home option, by 15 percentage points, which is an 85 percent increase from the baseline likelihood of 18 percent in the control group. Using administrative data from the survey company, the likelihood of showing up to perform the job outside the home increased by 11 percentage points, a 74 percent increase from the 15 percent likelihood in the control group. Both effects are statistically significant. Note that, as in most hiring contexts, here the firm interacts directly with the potential employee, not their spouse. The female employee may of course relay information to her spouse, especially if the information is aligned with her preferences, but that is beyond the control of the firm. By communicating directly with the women (as opposed to men, as in our first experiment), this additional experiment thus provides a “proof of concept” on how our conceptual contribution can be easily actionable in terms of policy to fix labor market distortions arising

from misperceived norms. Indeed, by increasing the share of women willing to work outside the home, this type of intervention could expand the set of jobs available to women.

We contribute to a growing literature on social norms in economics. This literature has focused mostly on the long-term *persistence* of cultural traits and norms (Fernández 2007, Giuliano 2007, Alesina, Giuliano, and Nunn 2013, Voigtländer and Voth 2015). We study how long-standing social norms can potentially change with the provision of information.⁹ We also contribute to a large literature on gender and labor markets (e.g., Bertrand 2011, Goldin 2014) by studying how social norms impact the share of women working outside the home, and our work is thus related to a growing literature on the effects of gender identity and norms on economic outcomes (see Akerlof and Kranton 2000; Fernández, Fogli, and Olivetti 2004; Fernández 2007; Eckel and Grossman 2008; Fernández and Fogli 2009; Jensen and Oster 2009; Dohmen et al. 2011; Alesina, Giuliano, and Nunn 2013; Baldiga 2014; Coffman 2014; Bertrand, Kamenica, and Pan 2015; Jayachandran 2015; Field et al. 2016; Bursztyn, Fujiwara, and Pallais 2017; Bernhardt et al. 2018; Bordalo et al. 2019; and McKelway 2019 for a discussion of the literature studying the role of social norms in explaining gender inequality in developing countries). Our paper relates to the work by Fernández (2013), which studies the role of cultural changes in explaining the large increases in married women's labor force participation over the last century in the United States. Our work adds to a growing literature on social image concerns in economics. Individuals' concerns about how they will be viewed by others has been shown to affect important decisions, from voting (DellaVigna et al. 2017, Perez-Truglia and Cruces 2017) to charitable donations (DellaVigna, List, and Malmendier 2012) to schooling choices (Bursztyn and Jensen 2015). We show that Saudi men's decisions to let their wives work outside the home are also affected by perceptions of the likelihood of judgment by others.

On the policy side, our results highlight how simple information provision might change perceptions of a society's opinions on important topics, and how this might eventually lead to changes in behavior. Conducting opinion polls and diffusing information about their findings could potentially be used to change important behaviors in some societies.

The remainder of this paper proceeds as follows. We discuss the experimental design of our main experiment and the underlying conceptual framework in Section I. In Section II, we present and interpret the results from that experiment. In Section III, we present the design and results from the national online survey and discuss evidence from the *Arab Barometer* survey and an additional survey on the frequency of discussion of the topic of WWOH. In Section IV, we present the design and results from the surveyor recruitment experiment. Section V concludes.

⁹Our paper also speaks to a recent theoretical literature on social norms (e.g., Bénabou and Tirole 2011, Ali and Bénabou 2016, and Acemoglu and Jackson 2017) by documenting how new information may lead to updates in perceptions of norms and fast changes in behavior.

I. Main Experiment: Design

A. Experimental Design

To organize thoughts and help guide our design, we present a simple model of labor supply and stigma in online Appendix A, following the intuition from Burszty, Egorov, and Fiorin (forthcoming). If married Saudi men believe that a large share of other men are opposed to WWOH and if they care to a great extent about their social image, they may end up not letting their wives work outside the home, despite the fact that most would prefer to allow their wives to work if the behavior were not observable. However, our model predicts that correcting perceptions about the opinions of others leads to drastic changes in the share of men willing to let their wives work outside the home.

Our experiment is comprised of two stages. We first conducted an experiment in the field to establish the effect of correcting participants' perceptions of the beliefs of others on a contemporaneous decision to sign up their wife for a job matching service. We then administered a follow-up survey three to five months later to collect information on longer-term labor supply outcomes.

Sample and Recruitment.—We partnered with the local branch of an international survey company and recruited 500 Saudi males between the ages of 18 and 35, living across Riyadh, Saudi Arabia, as participants in our main experiment. We targeted this age bracket primarily for logistical reasons in the sense that this population would be likely to use smartphones and also the mobile application/service that was to be offered in the experiment.¹⁰ Participants were required to have at least some college education as well as access to a smartphone.¹¹ We additionally restricted our recruitment to candidates who were married to ensure that participants would be able to make decisions regarding the labor market participation of their wives.

A recruiter database was used for initial contacts and further recruitment was conducted using a combination of snowball sampling and random street intercept. Participants were recruited from districts representing a range of socio-economic classes. Anticipating that contacts from districts with lower average incomes would be more responsive to the offered incentives, these districts were oversampled.¹²

Participants were organized into 17 experimental sessions of 30 participants each.¹³

Importantly, participants for each session were recruited from the same geographic area, so that participants in the same session shared a common social

¹⁰ An additional 120 participants were recruited for a pilot study consisting of 4 sessions which took place right before the start of the main experiment. The pilot study provided important logistical experience for survey facilitators and results were used to inform the final experimental design. Data from the pilot sessions are not included in the results presented.

¹¹ Access to a smartphone was required not only so that participants could take the survey on their own devices, but also so we can ask participants whether they would like to sign up for a job matching service which includes a mobile app.

¹² Participants were offered gift certificates with values ranging from SAR100–150 (\$26–40).

¹³ This means 510 total participants were expected, but only 500 completed the survey. Each session included two sequential survey links for each participant. The first part of the survey (corresponding to the first link) simply contained questions, the second part contained the informational conditions and the outcomes. The ten-subject attrition in the experiment is driven by participants who failed to activate the second link.

network. The average participant reported knowing 15 of the 29 other session participants. Sessions were held in conference rooms located in a Riyadh hotel, administered over the course of a week starting October 9, 2017 and ending October 13, 2017.

Main Experiment.—On their scheduled session date, participants arrived at the hotel. Upon arrival, participants were asked to provide their name, phone number, and email on a sign-in sheet before entering the designated session room. For each of the sessions, 5 rows of 6 chairs each were set up, with every other chair rotated to face the back of the room so that respondents' survey responses would not be seen by nearby participants. A survey facilitator instructed participants to sit quietly until the start of the survey, at which point all participants were instructed to begin the survey simultaneously. Participants were not allowed to ask further questions after the start of the survey.

The survey itself was administered using the online survey software Qualtrics and was implemented in two parts.¹⁴ In the first part of the survey, we collected demographic information and elicited participants' own opinions about a range of topics as well as their incentivized perceptions of others' beliefs. In the second part of the survey, we randomized participants to our information provision treatment and measure outcomes. At the start of the survey period, a survey link to the first part of the survey was provided on a board at the front of the room. Participants were instructed to navigate to the link and take the survey on their personal smartphones.¹⁵

Since WWOH may be a sensitive issue for respondents, maintaining anonymity of responses is an important focus of our experimental design. While the names and phone numbers of all participants were collected on the sign-in sheets, this identifying information was *not* collected in the survey itself. As a result, participants' names and full phone numbers are not associated with their individual survey responses at any point in the data collection or analysis process.¹⁶ Instead, participants were asked only to provide the last three digits of their phone number. We then used these trailing digits to randomize participants to treatment conditions, allowing us to recover a given participant's assigned treatment status using only their last three phone number digits.¹⁷ In addition, this also allows us to link participants' follow-up responses to their responses in the main experiment using only the combination of session number and these trailing digits.¹⁸ In sum, only the local surveyors collected a list with names and full phone numbers for each session.

¹⁴ All the scripts from the different interventions are included in online Appendix C.

¹⁵ While the majority of participants were able to take the survey successfully on their own devices, tablets were provided to those who encountered technical difficulties.

¹⁶ Participants were asked to provide an email address in order to receive a gift card reward for correctly guessing others' beliefs as well as the contact information of their wife if they choose to sign her up for a job matching service. All requests for identifying information occurred after the elicitation of private opinions and perceptions about the beliefs of others. Providing this contact information was also not required to complete the survey.

¹⁷ In particular, we electronically randomized treatment values to all possible combinations of three digits (000–999) before the start of the experiment. Participants were then assigned to the treatment condition corresponding to the treatment value pre-assigned to the last three digits of their phone number.

¹⁸ Provided that trailing digits identify participants one-to-one. Phone numbers are recorded on session specific sign-in sheets so that we only need to worry about matching digits within sessions; in practice we find that 12 within-session pairs have matching trailing digits. Since we have no way to match the follow-up responses of these respondents to their survey responses, we drop these respondents in the analysis of the follow-up calls, and these participants were not recontacted.

However, these surveyors were not able to link names to answers because they never had access to the data we collected through the online platform. Meanwhile, the researchers have access to the data, but do not have access to the participants' contact list—just the last three digits of their phone numbers. This anonymous design was chosen to facilitate the elicitation of honest opinions regarding WWOH.

In addition to using an anonymous online survey, we attempted to additionally reduce the scope for social desirability bias (SBS), by avoiding priming effects. The study was framed as a general labor market survey, with filler questions asking opinions on the employment insurance system, privileging Saudi nationals over foreigners into job vacancies, and the minimum wage level. In addition, no Western non-Arab was present during the intervention.¹⁹ In Section III we present results from an alternative approach to eliciting opinions regarding WWOH using a method that provides a higher degree of plausible deniability: the “list experiment.” The very similar findings there indicate that SBS is not likely to be an issue in our setting.

After collecting basic demographic information, participants were presented with a series of statements. For each statement, we asked the respondent whether they agreed or disagreed with the statement. We began with three statements regarding the labor market in general. These statements were as follows:

- *In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.*
- *In my opinion, the current unemployment insurance system (Haafez) is good for the economy.*
- *In my opinion, the minimum wage for Saudis (SAR 3,000) should be kept at its current level.*

We then presented two statements regarding the participation of women in the labor force:

- *In my opinion, women should be allowed to work outside of the home.*
- *In my opinion, a woman should have the right to work in semi-segregated environments.*

For each of the statements regarding WWOH as well as the statement about the minimum wage, participants were asked to estimate how many of the other 29 participants they expected to agree with the statement.²⁰ To incentivize participants to

¹⁹SBS would be an important issue if it leads to an upward bias in reported pro-WWOH opinions. We note that it is unclear in what direction the bias would go: if participants are afraid of stigma associated with being pro-WWOH in the case that their neighbors discover their true opinion, we would be *underestimating* the true level of support.

²⁰We asked two questions regarding WWOH, with the first one being our primary question and the second one added for robustness. In general, the object of interest was whether women should or should not be allowed (i.e., have the right) to work outside of home under current labor market conditions in Saudi Arabia. Conceptually, this right would arguably apply to any subset of circumstances. By contrast, we were not aiming to ask whether it is *optimal* for women to work outside of home under a particular circumstance, which would require qualifiers. To address the possibility that subjects misinterpret the first question, perhaps because it leaves it open whether men and women are fully integrated in workplaces (which is not the current situation), the second question refers to what is done in practice in Saudi Arabia today. Semi-segregated workplaces have a separate section for women and men, which is required by law. We also replace the word “allowed” with “have the right,” to make sure there is no ambi-

estimate the beliefs of others accurately, participants were told that the respondent with the closest guess would receive a \$20 Amazon gift card code (distributed by email after the end of the session).

This concluded the first part of the survey. Upon completing this part of the survey, participants were instructed to wait for the facilitator's permission before continuing. The session facilitator was instructed to closely monitor the progress of the survey responses through the online platform. After it had been determined that all participants had completed the first part of the survey, participants were instructed to continue on to the second part of the survey by following a link provided on the end page of the first survey.

In the second part of the survey, one-half the participants were randomly assigned to a treatment condition. This randomization was conducted at the individual level, based on the last three digits of the respondents' phone number (provided in the first part of the survey). In the treatment condition, participants were given feedback about the responses of the other session participants to the two statements about female labor participation. In particular, we provided treatment participants with charts embedded in the survey interface showing the proportion of respondents who reported agreeing and disagreeing with each statement in the first part of the survey. Participants assigned to the control group received no information.

Outcomes.—All participants were then asked to read a short passage about a Saudi start-up which provides an online platform aiming to connect job-seeking Saudi women with employers. The start-up's platform lists thousands of job opportunities for Saudi women, matching workers to jobs based on skills and interests. The informational passage contained basic information about the online platform as well as the company's outreach and mentorship initiatives.

We then asked respondents to make an incentivized choice between receiving a \$5 Amazon gift card and the opportunity to sign up their wives for access to the company's platform and services. An important service in the case of sign up is a weekly email to be received listing a number of links to descriptions and application forms for postings of jobs outside the home for women in the participants' areas. Participants who chose to sign up for the service were subsequently asked to provide the contact information of their wife.²¹ Note that since a participant's wife's eventual employment status is observable to peers, the observability of the sign up choice itself does not matter independently, so we chose not to randomize the observability of the sign up decision.

If participants who privately support WWOH underestimate the level of support for WWOH among others, they may incorrectly expect that signing up their wives for the job matching service is stigmatized. If this is indeed the case, participants who receive information correcting their beliefs about others' support for WWOH (treatment) should exhibit a higher sign up rate than participants who receive no information (control).

guity. As we will show below, we find very similar results regardless of the wording, indicating that we successfully measure the object of interest using our primary question.

²¹In the results we present, as pre-registered, those who do not provide their wife's name and number (even if initially choosing the sign up option) are treated as having chosen not to sign up. Results are almost identical if we include those as effectively signing up.

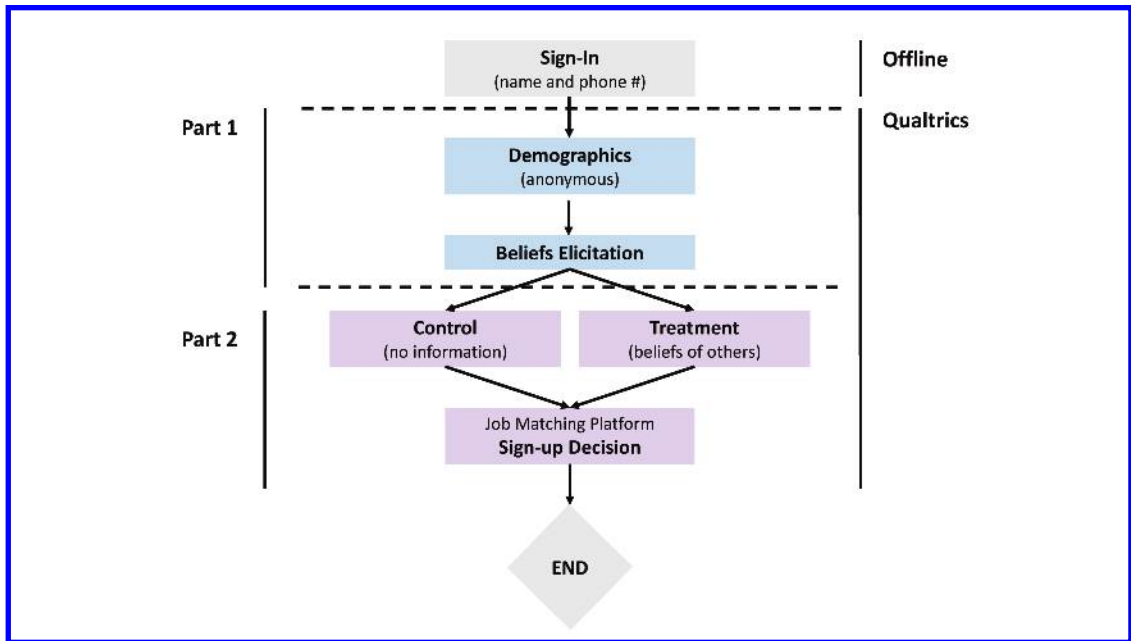


FIGURE 1. EXPERIMENTAL DESIGN: MAIN EXPERIMENT

Notes: Experimental design of the main experiment. Dashed lines indicate points in the survey flow where participants were instructed by the session facilitator and/or the survey instructions to wait for all session participants to be ready to start the next section before proceeding.

At the end of our survey, we ask participants to estimate the percent of private sector firms in Saudi Arabia that have semi-segregated work environments as a way to proxy for perceived market demand for female labor outside the home. Observing that our information provision has a positive effect on the expected percentage of firms with semi-segregated environments would provide evidence consistent with perceived female labor demand being a function of perceived social norms regarding WWOH. We use this information to assess the extent to which the decision to sign up is driven by changes in perceptions about labor demand—a different mechanism that is potentially relevant.

Figure 1 summarizes the design of the main experiment.

Follow-Up Calls: Longer-Term Outcomes.—We conducted follow-up calls three to five months after the conclusion of the main experiment in order to collect longer-term labor market outcomes.²² We partnered with the same survey company to re-contact participants and conduct the follow-up survey.

Participants were contacted using the phone numbers collected on the session sign-in sheets during the main experiment. Participants who were not reached initially were called several times. At the beginning of each call, the surveyor reminded the respondent of his participation in the main experiment and provided further information about the follow-up call. Participants who consented to participate in

²²Follow-up calls began on January 10, 2018 and ended March 6, 2018.

the follow-up were offered a payment of up to SAR 35 in phone credit for successful completion of the survey.

As noted above, we link the data collected in the follow-up call to individual survey data in the main experiment, using the last three digits of the respondent's phone number. Twelve pairs of respondents in the same experimental session provided identical trailing digits. We drop these participants from the subsequent follow-up analysis since we cannot uniquely match them to their survey responses. Surveyors were able to recontact 82 percent of the original experimental subjects with unique trailing digits, and 98 percent of recontacted participants completed the survey. If one takes into account the twelve pairs of respondents with identical trailing digits at the session level (that is, 4.8 percent of the original sample), the total attrition rate is therefore 24 percent. We find no evidence of selective attrition overall, or within initial conditions (see online Appendix Tables B1 and B2), both based on observables and on the sign up decision in the original experiment. We note that it is possible that some respondents changed their cell phone numbers in the four months between the experiment and the follow-up survey, which could have contributed to the attrition rate (a number of numbers called were disconnected, which could have resulted from changes in telephone service or, alternatively, to participants providing incorrect numbers in during the experiment). In Section IIA we provide Lee (2009) attrition bounds for our treatment effects.

In each follow-up call, we collected a variety of outcomes related to the wife's labor supply outcomes. We asked whether the wife was currently employed and if so, whether this job was outside the home. Jobs outside the home are our outcome of interest since the information intervention focuses on norms about working outside the home, and the job matching service focuses on access to jobs of this kind. We also asked whether the wife was employed three to four months ago (depending on the follow-up survey data), i.e., at the time of the main experiment. We did this for two reasons. First, since we had already asked whether the wife was employed during the main experiment itself, asking the question again in the follow-up allowed us to compare responses to the two questions as a sanity check for internal consistency. Reassuringly, 94 percent of respondents respond consistently, giving the same answer to the question in both instances. Second, we were initially surprised that almost two-thirds (65 percent) of respondents in the main experiment reported that their wife was employed. This high proportion led us to question whether some of the respondents' wives were in fact working from home. We used the follow-up survey as an opportunity to ask whether respondents' wives were employed *outside* of home. We find that while 61 percent of the respondents in the follow-up survey reported that their wives were working three months before (i.e., around the time of the original experiment), only 8 percent of them reported that their wives were then working *outside the home*.²³ Participants were also given the option to answer an open-ended question asking for their wives' exact occupation. However, only a small share of participants provided an answer to this question, so we interpret the numbers with caution. The most common occupation categories for wives

²³ In the larger, online national survey we conducted later, 43 percent of participants reported that their wife was currently working, while only 4 percent reported that their wife was working outside the home.

working from home were customer service: via phone calls (51 percent) and data entry (20 percent).

We then asked whether their wife participated in any job search activities in the last three to four months. In particular, we collected information on whether their wife applied for a job during this period, interviewed for a job during this period, and whether she had any interviews currently scheduled. For all questions, affirmative answers were followed by a question asking whether the job of interest was outside the home. Next, we considered whether the information intervention in the original experiment may have had spillover effects on related behavioral outcomes. Taking advantage of the recent announcements lifting the ban on women's right to drive, we posed a hypothetical question asking participants if they would be willing to sign up their wife for driving lessons if given the chance. Finally, we checked for persistence in the correction of perceptions of others' beliefs induced by the information provision treatment. In particular, we asked participants to guess how many people out of 30 randomly selected residents of their neighborhood would agree with each of the statements presented in the main experiment. Finding systematic differences in perceptions about norms between the treatment group and the control group would suggest that the information provision treatment had a persistent effect on participants' beliefs about others. Since participants were recruited from the same neighborhoods within sessions, information spillovers from treatment to control participants could have washed out differential persistence in perceived norms.

Due to the number of outcomes we collect, we address the issue of multiple hypothesis testing by constructing an index pooling all of our collected outcomes following Kling, Liebman, and Katz (2007). We discuss the construction of this index in Section IIA. We also adjust *p*-values for multiple hypothesis testing following the procedure developed by List, Shaikh, and Xu (2019).

II. Main Experiment: Results

A. Main Results

Misperceived Social Norms.—We start by describing the measured misperception about social norms relating to WWOH. The average (median) guess is that 63 percent (67 percent) of other session participants agree with the pro-WWOH statement. The average level of agreement across all sessions is 87 percent, a number larger than the guesses of close to 70 percent of participants. Of course, the proper comparison is between a participant's guess and the share agreeing *in his session*. In Figure 2, we calculate the wedge between the participant's belief about the opinions of other session subjects and the actual opinions of all session participants, and then plot the distribution of these wedges across all sessions. We find that 72 percent of participants strictly underestimate the support for WWOH among session participants (noting that these are individuals recruited from the same neighborhoods), with an average wedge of 24 percentage points.²⁴

²⁴Online Appendix Figure B1 displays very similar findings for the second question relating to WWOH (semi-segregated environments).

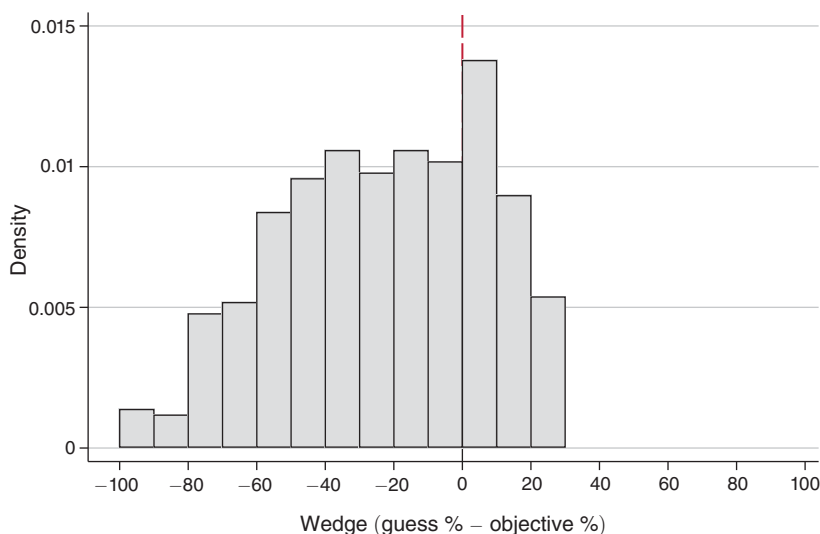


FIGURE 2. WEDGES IN PERCEPTIONS OF OTHERS' BELIEFS: MAIN EXPERIMENT

Notes: The distribution of wedges in perceptions about the beliefs of others regarding whether women should be able to work outside the home. Wedges calculated as *(the respondent's guess about the percent of session participants agreeing with the statement) - (the true percent of session participants agreeing with the statement)*.

To further examine what is underlying these guesses, we also asked people to assess, on a scale of 1–5, how confident they were with their guesses. We also asked how many other people they knew in the room. In online Appendix Figure B2, we show that more accurate guesses are significantly correlated with higher levels of confidence in the guess, and that knowing more people in the room is significantly correlated with both higher levels of confidence in the guess and with more accurate guesses. These are just suggestive correlations, but they suggest that men partially learn about the acceptability of female labor supply behavior from the people around them. We further investigate this hypothesis in Section IIIB.

The information collected in the follow-up survey on whether wives were working outside the home at the time of the original experiment allows us to conduct another validity check: 100 percent of participants whose wife was working outside the home reported supporting WWOH outside the home. We also note that participants supportive of WWOH outside the home (the vast majority of participants in the sample) have a significantly higher perception of support by other session subjects than participants opposed to WWOH outside the home (average guesses of 18.7 versus 13.4, respectively, with $p = 0.000$).

Main Results (1): Job Matching Service Sign Up Rates.— We now turn to our experimental intervention correcting beliefs about the opinions of other participants. Table 1 provides evidence that individual characteristics are balanced across the two experimental conditions, confirming that the randomization was successful.

Figure 3 displays our main findings. In the control condition, 23.5 percent of participants prefer to sign up their wives for the job service as opposed to taking the gift card payment. In the group receiving feedback on the responses of other participants

TABLE 1—SUMMARY STATISTICS: MAIN EXPERIMENT

	All	Control	Treatment
Observations	500	247	253
Age	24.78 (4.21)	24.64 (3.99)	24.91 (4.41)
Number of children	1.71 (1.72)	1.64 (1.70)	1.77 (1.74)
College degree (%)	56.20	55.06	57.31
Employed (%)	86.60	87.45	85.77
Wife employed (%)	65.20	65.59	64.82
Wife working outside the home (% retrospective follow-up)	8.40	7.89	8.90
Other participants known (%)	51.19 (38.24)	49.68 (38.60)	52.66 (37.92)
Other participants with mutual friends (%)	38.64 (34.94)	37.62 (34.62)	39.63 (35.29)

Notes: Baseline summary statistics of respondent characteristics in the main experiment. Summary statistics on wife working outside the home are from the self-reported labor supply status of participants' wives three months before the experiment in the follow-up survey (sample size = 381).

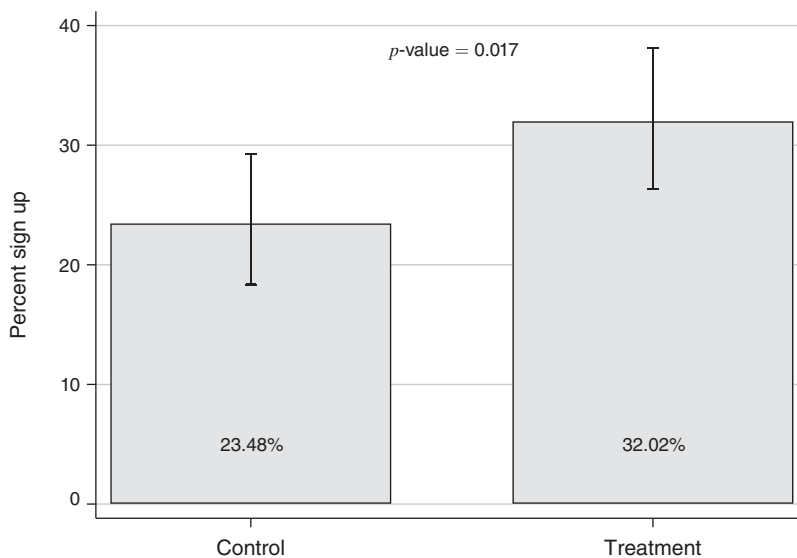


FIGURE 3. JOB MATCHING SERVICE SIGN UP: MAIN EXPERIMENT

Note: Job matching service sign up rates for respondents in the main experiment; 95 percent binomial proportion confidence intervals; p -value calculated from testing for equality of proportions.

to the questions on support for WWOH, the sign up rate goes up to 32.0 percent, a 36.4 percent increase (the p -value of a t -test of equality is 0.017). Table 2 displays the findings in regression format: column 1 replicates the findings from Figure 3. In column 2, we add session fixed effects. For robustness and potential efficiency gains, column 3 controls for baseline beliefs and column 4 includes additional individual covariates. Results are unchanged across specifications. In this table and all

TABLE 2—WEDGES IN PERCEPTIONS OF OTHERS' BELIEFS: MAIN EXPERIMENT

	(1)	(2)	(3)	(4)
Treatment (β)	0.0853 (0.0399)	0.0876 (0.0399)	0.0899 (0.0395)	0.0898 (0.0404)
Constant	0.235 (0.0270)	0.132 (0.0739)	−0.156 (0.0970)	−0.131 (0.186)
<i>Inference robustness (β)</i>				
<i>p</i> -value: robust SE	0.033	0.028	0.023	0.027
<i>p</i> -value: wild bootstrap	0.018	0.013	0.013	0.008
<i>p</i> -value: permutation test	0.036	0.038	0.024	0.025
Session fixed effects		✓	✓	✓
Baseline beliefs			✓	✓
Controls				✓
Observations	500	500	500	491
R^2	0.00907	0.0696	0.108	0.116

Notes: Column 1 reports estimates from an OLS regression of an indicator for whether the respondent signed their wife up for the job matching service on a treatment dummy. Column 2 includes session fixed effects. Column 3 controls for the respondent's own opinions and perceptions of others' beliefs at baseline regarding the labor market statements described in the main experiment. Column 4 adds socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parentheses. Reported *p*-values for wild bootstrap and permutation tests derived from running 1,000 replications in each case. Wild bootstrap clustered at the session level.

tables which follow, in addition to *p*-values from robust standard errors, we also present *p*-values from wild cluster bootstrap standard errors (where we cluster at the session level) and permutation tests for relevant coefficients.²⁵

Main Results (2): Longer-Term Outcomes.—We now evaluate the longer-term effects of the information intervention (three to five months later). Figure 4 displays the results. We first present the impact of the original treatment on three self-reported labor-supply outcomes: whether or not the participant's wife applied for a job outside the home since the original intervention, whether or not she interviewed for a job outside the home during this period and whether or not she is currently employed outside the home. Across all outcomes, we see increases stemming from the original treatment, though this difference is significant only for the first two outcomes. The percentage of wives who applied for a job outside the home during the time frame of interest goes up from 5.8 percent to 16.2 percent, a 180 percent increase (the *p*-value of a *t*-test of equality is 0.001). Regarding interviews for this type of job, the share increases from 1.1 percent to 5.8 percent (*p*-value = 0.006). Rates of employment outside the home go up from 7.4 percent to 9.4 percent (*p*-value = 0.235).

We also examine whether the information provided might lead to changes in other behaviors related to women's rights in Saudi Arabia. Figure 4 shows that the share of husbands who report in a hypothetical question that they would sign their wives up for driving lessons goes up from 65.2 percent to 76.4 percent (*p*-value = 0.008). This suggests that the effects of the information intervention

²⁵ We use wild cluster bootstrap standard errors because of the small number of clusters (Cameron, Gelbach, and Miller 2008). We use permutation tests due to potentially small sample sizes.

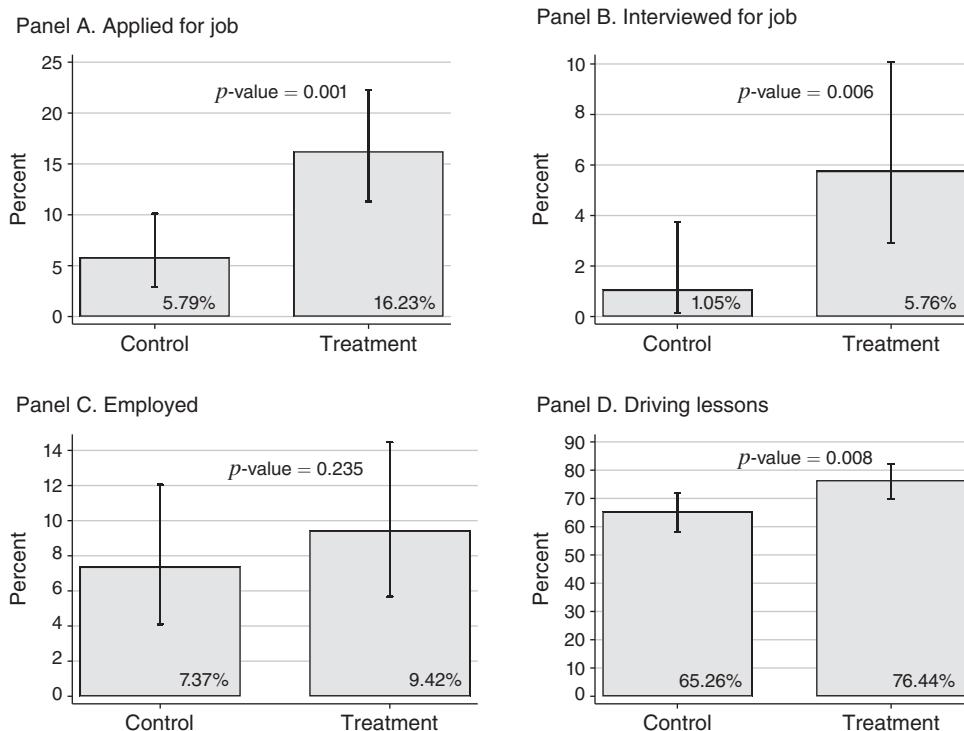


FIGURE 4. LONG-TERM LABOR SUPPLY OUTCOMES: FOLLOW-UP

Notes: Self-reported labor supply outcomes of participants' wives in the follow-up survey; 95 percent binomial proportion confidence intervals; p -value calculated from testing for equality of proportions. Panels A–C refer exclusively to job opportunities outside the home.

extend beyond behaviors narrowly associated with WWOH outside the home, and may act more generally with respect to norms about women's rights in Saudi Arabia.

We next assess whether the original treatment produced a persistent change in perceived norms (and whether spillovers via subsequent communication with the control group undid the original intervention). To that end, we asked participants to provide an unincentivized guess on the number of 30 randomly selected people *from their neighborhood* they think would agree with the same pro-WWOH statements from the original experiment. We find evidence of a persistent change in perceptions by treated relative to control participants: the average guessed share for “working outside the home” increases from 43.9 percent to 56.6 percent (p -value = 0.000). These findings also suggest that the intervention updated beliefs about neighbor's opinions in general, not just regarding responses by other session participants.²⁶

To deal with the potential issue of multiple hypothesis testing, we follow Kling, Liebman, and Katz (2007), and construct an index pooling all six of our used

²⁶In online Appendix Figure B3 we show that there was no treatment effect on the filler question regarding the opinions of neighbors on the current level of the minimum wage. Although there is a theoretical possibility that subjects could have updated their views on the opinions of others regarding issues not directly related to gender issues in the labor market—but labor market issues more broadly—we find no evidence to that effect.

TABLE 3—EFFECT OF TREATMENT ON LABOR SUPPLY OUTCOMES: FOLLOW-UP

	K-L-K index	Employed	Applied	Interviewed	Driving lessons	Beliefs about neighbors
<i>Panel A. No controls</i>						
Treatment (β)	0.249 (0.0495)	0.0206 (0.0285)	0.104 (0.0317)	0.0471 (0.0185)	0.112 (0.0463)	0.128 (0.0267)
Constant	-0.119 (0.0272)	0.0737 (0.0190)	0.0579 (0.0170)	0.0105 (0.00742)	0.653 (0.0346)	0.439 (0.0195)
<i>Inference robustness (β)</i>						
<i>p</i> -value: robust SE	0.000	0.471	0.001	0.011	0.016	0.000
<i>p</i> -value: wild bootstrap	0.000	0.543	0.032	0.010	0.019	0.001
<i>p</i> -value: permutation test	0.000	0.593	0.002	0.017	0.009	0.000
<i>p</i> -value: L-S-X MHT correction	—	0.484	0.003	0.050	0.039	0.000
<i>Lee attrition bounds</i>						
Lower bound	0.238 (0.0576)	0.014 (0.0288)	0.101 (0.0314)	0.046 (0.0181)	0.092 (0.0564)	0.119 (0.0350)
Upper bound	0.276 (0.0592)	0.033 (0.0532)	0.120 (0.0555)	0.056 (0.0166)	0.111 (0.0500)	0.138 (0.0392)
Observations	381	381	381	381	381	381
R^2	0.0626	0.00137	0.0278	0.0168	0.0151	0.0571
<i>Panel B. Session fixed effects, baseline beliefs, and socioeconomic controls</i>						
Treatment (β)	0.276 (0.0528)	0.0223 (0.0282)	0.115 (0.0332)	0.0469 (0.0182)	0.125 (0.0475)	0.144 (0.0268)
Constant	-0.141 (0.210)	0.0545 (0.137)	-0.0262 (0.146)	0.0222 (0.0795)	0.882 (0.225)	0.310 (0.114)
<i>Inference robustness (β)</i>						
<i>p</i> -value: robust SE	0.000	0.429	0.001	0.010	0.009	0.000
<i>p</i> -value: wild bootstrap	0.000	0.438	0.026	0.001	0.022	0.002
<i>p</i> -value: permutation test	0.000	0.423	0.000	0.006	0.007	0.000
Observations	375	375	375	375	375	375
R^2	0.164	0.0947	0.166	0.117	0.128	0.229

Notes: Each column reports estimates from OLS regression of the outcome indicated by the column on a treatment dummy. The Kling-Liebmman-Katz index, defined in the text, is constructed from all six tested outcomes including those not reported in the table. Panel A includes no controls while Panel B adds session fixed effects, baseline beliefs (own opinions and perceptions of others' beliefs) and socioeconomic controls for age, education, employment status (of both respondent and wife), number of children, and the share of people in the session room the respondent reported knowing and having mutual friends with. In panel A, L-S-X MHT correction refers to the multiple hypothesis testing procedure presented in List, Shaikh, and Xu (2019). Robust standard errors reported in parenthesis. Reported *p*-values for wild bootstrap and permutation tests derived from running 1,000 replications in each case. Wild bootstrap clustered at the (main-experiment) session level.

outcomes.²⁷ The index is $(1/6)\sum_{k=1}^6(k - q_k)/\sigma_k$, where k indexes an outcome, and q_k and σ_k are the mean and standard deviation of that outcome for participants originally in the control group. In addition, we display *p*-values adjusting for multiple hypothesis testing following the procedure developed by List, Shaikh, and Xu (2019). Finally, we provide Lee (2009) attrition bounds for our treatment effects.

Table 3 presents the results in regression format, with and without controls.

We can further examine the longer-term effects of the information provided on perceptions of social norms. Online Appendix Figure B4 provides visual evidence

²⁷The index here includes six outcomes (interviews scheduled is included in the index but not in the tables since only 1 person had an interview scheduled). Results are unchanged if we include as a seventh outcome guesses about opinions regarding the statement on semi-segregated environments, since the answers were very correlated for both types of guesses.

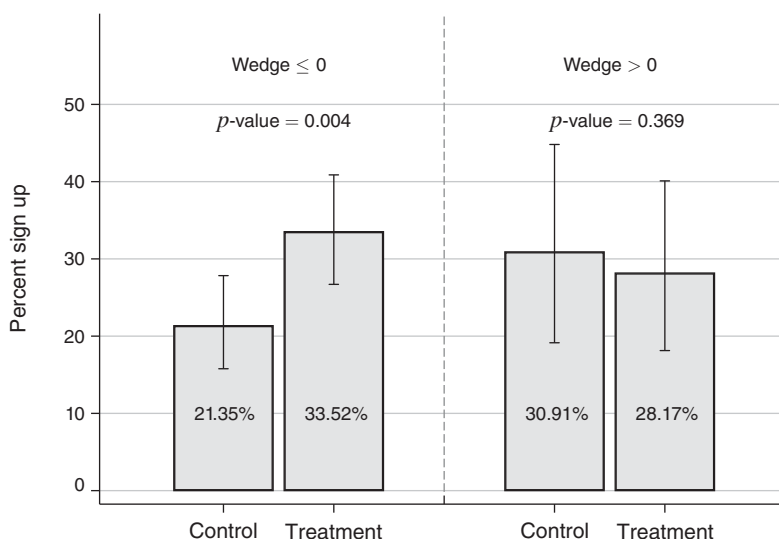


FIGURE 5. JOB MATCHING SERVICE SIGN UP, HETEROGENEITY BY WEDGE: MAIN EXPERIMENT

Notes: Job matching service sign up rates for respondents with non-positive and positive wedges in perceptions about the beliefs of others regarding whether women should be able to work outside the home. Wedges calculated as (the respondent's guess about the percent of session participants agreeing with the statement) – (the true percent of session participants agreeing with the statement); 95 percent binomial proportion confidence intervals. p -value calculated from testing for equality of proportions.

that for control group participants, higher beliefs about the share of pro-WWOH session participants are associated with higher beliefs about the share of pro-WWOH neighbors in the follow-up survey (the effect is significant at the 1 percent level). The figure also shows that the levels of beliefs in the follow-up survey are higher for treated participants, and that the relationship between prior and posterior beliefs is (significantly) flatter for treated participants. These results suggest that the treatment indeed updated priors about neighbors' opinions and that treated participants used the signal from the treatment to override their priors. Online Appendix Table B3 displays the results in regression format with session fixed effects and controls.

B. Heterogeneity by Direction of Priors' Update

We move beyond the average effects of the information provision and now analyze heterogeneous effects by the direction (and size) of priors' updating. Figure 5 provides the raw averages and Table 4 displays regression results. For those who originally underestimated the true share of session participants agreeing with the statement regarding women's right to work outside the home, we observe a large increase in sign up rates for the job matching service when information is provided: the share signing up increases from 21.4 percent in the control group to 33.5 percent in the treatment group (a 57 percent increase, p -value = 0.004). For those who originally overestimated the true share of session subjects agreeing with the statement, the baseline sign up rate is higher in the control group (30.9 percent), which is by itself consistent with the hypothesis of a role of

TABLE 4—JOB MATCHING SERVICE SIGN UP: HETEROGENEITY BY WEDGE: MAIN EXPERIMENT

	Wedge ≤ 0				Wedge > 0			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment (β)	0.122 (0.0459)	0.122 (0.0457)	0.127 (0.0450)	0.125 (0.0463)	−0.0274 (0.0827)	−0.0383 (0.0883)	−0.0391 (0.0919)	−0.0252 (0.114)
Constant	0.214 (0.0297)	0.0511 (0.0710)	−0.202 (0.117)	−0.251 (0.204)	0.309 (0.0628)	0.315 (0.162)	0.711 (0.659)	0.946 (0.882)
<i>Inference robustness (β)</i>								
<i>p</i> -value: robust SE	0.008	0.008	0.005	0.007	0.741	0.665	0.671	0.825
<i>p</i> -value: wild bootstrap	0.006	0.011	0.004	0.002	0.735	0.689	0.728	0.854
<i>p</i> -value: permutation test	0.014	0.009	0.007	0.010	0.850	0.644	0.651	0.826
Session fixed effects		✓	✓	✓		✓	✓	✓
Baseline beliefs			✓	✓			✓	✓
Controls				✓				✓
Observations	374	374	374	365	126	126	126	126
R^2	0.0186	0.0943	0.135	0.150	0.000890	0.136	0.221	0.271

Notes: Wedge for the given statement (whether women should be able to work outside the home) calculated as *(the respondent's guess about the number of session participants agreeing with the statement) – (the true number of session participants agreeing with the statement)*. Column 1 reports OLS estimates from regressing an indicator for whether the respondent signed their wife up for the job matching service on a treatment dummy restricted to respondents with non-positive wedge (those underestimating the progressivism of the other participants). Column 5 reports estimates from the same specification restricted instead to respondents with positive wedge. Columns 2 and 6 include session fixed effects. Columns 3 and 7 additionally control for respondents' opinions and perceptions of others' beliefs at baseline regarding the labor market statements described in the main experiment. Columns 4 and 8 add socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parentheses. Reported *p*-values for wild bootstrap and permutation tests derived from running 1,000 replications in each case. Wild bootstrap clustered at the session level.

perceived social norms affecting labor supply decisions for their wives. The effect of the intervention is no longer significant for that group, and is in fact directionally negative (albeit small). We note that the interaction term in the regression including the whole sample (that is, the treatment interacted with a dummy on the sign of the wedge) is not significant at conventional levels (the *p*-value of the interaction in the regression without additional covariates is 0.115, while it is 0.204 for the full specification).²⁸ Interestingly, the sign up rate among control participants with positive wedges is very similar (and not statistically different) from the sign up rate in the treatment group, which provides correlational evidence from the control group consistent with our hypothesis that perceptions of social norms matter for the sign up decision. The evidence on heterogeneous updating for the longer-term outcomes is less conclusive, and is shown in online Appendix Tables B4 (without controls) and B5 (full specification).²⁹

²⁸ Also note that, as discussed before, participants with larger negative wedges are more likely to be personally opposed to WWOH, which in principle would reduce the scope for an effect of the information provision to them. When we restrict the regression to those who reported supporting WWOH, the interaction term yields *p*-values of 0.099 (no additional controls) and 0.209 (full specification).

²⁹ We find similar treatment effects on sign up rates for participants with fewer or more social connections in the original experimental session. We note though that a priori, it is unclear whether the effects should be stronger or weaker according to the number of social connections: the treatment might be less informative because socially connected participants might already have better guesses about the opinions of others, but those participants might potentially care more about the opinions of the other subjects, thus strengthening the treatment.

TABLE 5—EFFECT OF BELIEF UPDATE ON JOB MATCHING SERVICE
SIGN UP: MAIN EXPERIMENT

	(1)	(2)	(3)
Update ($-\text{Wedge} \times \mathbf{1}_{\text{Treatment}}; \beta$)	0.00758 (0.00334)	0.00648 (0.00342)	0.00682 (0.00352)
Constant	−0.0255 (0.0716)	−0.116 (0.0979)	−0.111 (0.182)
<i>Inference robustness</i> (β)			
<i>p</i> -value: robust SE	0.023	0.058	0.053
<i>p</i> -value: wild bootstrap	0.008	0.015	0.026
<i>p</i> -value: permutation test	0.028	0.069	0.069
Baseline beliefs and confidence	✓	✓	✓
Session fixed effects		✓	✓
Controls			✓
Observations	500	500	491
R^2	0.0270	0.0798	0.0926

Notes: Column 1 reports estimates from an OLS regression on the update in belief about the support of others induced by the information treatment (for women working outside the home), controlling for own baseline opinions, perceptions of others' beliefs, and confidence in beliefs about others (for the question asking whether women should be able to work outside the home). The update measure is given by minus the wedge for those in the treatment condition and equals 0 for those in the control condition. Column 2 includes session fixed effects. Column 3 additionally includes socioeconomic controls for age, education, employment status (of both respondent and wife), number of children, and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parentheses. Reported *p*-values for wild bootstrap and permutation tests derived from running 1,000 replications in each case. Wild bootstrap clustered at the session level.

If one assumes that treated participants use the numbers given by the information provision as their posterior beliefs about the other session subjects, we can compute a *continuous* measure of belief update during the experiment. For control participants, this update is equal to zero. For treatment participants, it is equal to the negative of their “wedge.” We can then evaluate the impact of the *size of the prior updating* on the sign up decision, as displayed in Table 5. Note that unlike in the other tables, we control for participants' own views on WWOH, their beliefs about other session subjects, and their confidence in their own guesses, since conceptually all these variables should matter for a given prior updating level. We find that higher levels of prior updating lead to significantly higher sign up rates. For example, using the specification in column 1, a positive update in prior corresponding to one standard deviation of the wedge in the treatment group causally leads to a 6.9 percentage point increase in the job matching service sign up rate. In online Appendix Table B6, we provide suggestive evidence that higher levels of updating also lead to larger changes in the longer-term outcomes.

C. Interpreting the Results

Understanding the Longer-Term Effects.—It is difficult to separate the extent to which the longer-term effects are driven by the higher rate of access to the job service versus a persistent change in perceptions of the stigma associated with WWOH. The fact that we observe a persistent treatment effect on beliefs about neighbors

measured in the follow-up survey and on the (reported) willingness to sign up one's wife for driving lessons (that is, a different, related outcome) suggests that the longer-term effects on job search behavior might not be entirely mechanically driven by the job matching service alone.

One might wonder if there could have been informational spillovers between treatment and control participants. These participants often reported knowing each other and could have discussed the information given to them after the main experiment. This is particularly related to our hypothesis; if perceived social sanctions associated with WWOH have been reduced, some participants would be more open to discussing the topic with their neighbors or friends. While we observe persistent differences in beliefs about others in the follow-up survey, we have no way of knowing if the differences between control and treatment groups would have been even larger in the absence of potential spillovers.³⁰

Interestingly, we find significant effects for outcomes (applying and interviewing for jobs outside the home) that are not easily observable (when compared to working outside the home itself). It is possible that the difference in beliefs about others between the two conditions would have been smaller if control participants started observing more women working outside the home. Unfortunately, we do not have the data to test this hypothesis.³¹

Signaling Labor Demand.—The channel we have in mind to explain our results is an update in the perceived social sanctions associated with having a wife working outside the home. However, updating one's beliefs about what others think regarding WWOH might also turn on a potential additional channel, since this update may also lead to learning about the labor demand side. Subjects may think “If so many people, in fact, support WWOH, then there are probably many firms willing to hire women for jobs outside the home.” This is an interesting alternative mechanism and could explain our findings if (i) inducing positive updates in perceptions about other participants' opinions leads to positive updates in perceptions about the availability of jobs for women, and (ii) this second update matters for the sign up decision (i.e., participants originally believed labor demand was a constraint). In online Appendix Table B7, we find that the treatment leads to a small (and generally not-significant) increase in participants' beliefs about the percentage of private-sector firms in Saudi Arabia with semi-segregated environments. In online Appendix Table B8, we show that once we include covariates, there is no correlation in the control group between participants' perception of firms' willingness to hire women and participants'

³⁰We find similar treatment effects on longer-term outcomes for participants with fewer or more social connections in the original experimental session, suggesting that informational spillovers alone might not have been enough to induce more job search among control participants' wives. Possible reasons for this pattern include (i) the other, known participants were not very close friends but their opinions are still relevant signals about the relevant peer group in the neighborhood; (ii) perhaps one has no economic incentives to tell one's friends until one's wife gets a job because communicating to others would increase competition for the same jobs; (iii) if treated participants now think that WWOH is socially desirable and a symbol of status, they may want to be the first to have an employed wife (to distinguish themselves), or derive utility from being the “leader” and having others follow them.

³¹An interesting question is the extent to which our results are driven by households without children—given that childcare could be an additional obstacle for women to take a job outside the home. In online Appendix Figures B5, B6, and B7, we present our results in terms of sign up rates, longer-term outcomes, and the outcomes for the recruitment experiment, separately for women with or without children. We do not see any striking pattern of heterogeneity on that dimension, though for some of the outcomes the effects seem weaker for households with children.

decision to sign up for the job matching service. We therefore believe that this alternative mechanism is not driving our findings.

III. National Surveys and Discussion of Misperceptions Drivers

A. *First National Survey*

We now examine results from an anonymous, online survey with a larger, national sample of about 1,500 married Saudi men that used the same platform and had the same layout as the original experiment. The goal of this additional survey is twofold. First, we assess the external validity of the finding that most Saudi men privately support WWOH while failing to understand that others do. Second, we examine whether social desirability bias/experimenter demand effects could have been a driver of the misperception finding in the main experiment.

Survey Design.—The sample of survey takers was recruited through the same survey company as before and the (visually identical) survey was administered online through Qualtrics. As in the main experiment, we restricted the survey to married Saudi males between the ages of 18 and 35. The final sample of participants was designed to be a nationally representative sample of the targeted demographic category, enabling us to provide evidence of the external validity of our main finding.

After signing an online consent form, participants were asked the same set of demographic questions as in the main experiment. Then, we departed from the design of the main experiment by implementing a list experiment to introduce plausible deniability to our elicitation of individual beliefs. In a list experiment, participants are randomly assigned to either a control group or to a treatment group. In the control group, participants are presented with a list of statements or policy positions that are contentious but not stigmatized. In the treatment group, participants are presented with an identical list of items but also an additional, potentially stigmatized item for which the experimenter would like to elicit beliefs. Participants in both conditions are then asked to indicate how many of the statements they agree with. The true degree of support for the item of interest at the sample level can then be inferred by comparing the average number of agreements in the treatment group with the average number of agreements in the control group.

This design allows the experimenter to ask participants only *how many* of the statements or policy positions in a given list they agree with rather than needing to know which items they support in particular. In our case, as long as participants do not disagree with all the non-WWOH statements, the list experiment provides plausible deniability to those who do not support WWOH and might otherwise be affected by social desirability bias.

We operationalized the list experiment in our online survey as follows. Participants in the control group were presented with a list of three general statements about the labor market (the same statements used in the main experiment).³² These statements were chosen to be contentious but *not* stigmatized:

³²The ordering of the statements in the list was randomized for both the treatment and the control groups.

- *In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.*
- *In my opinion, the current unemployment insurance system (Haafez) is good for the economy.*
- *In my opinion, the minimum wage for Saudis (SAR 3,000) should be kept at its current level.*

Participants were asked to read all three statements carefully and to indicate the number of statements (from 0 to 3) that they agreed with. Note that we did not ask participants which of the statements they agreed with, only how many.

In the treatment group, participants were presented with the same list of statements with the single addition of the statement of interest regarding WWOH that *is* potentially stigmatized:

- *In my opinion, women should be allowed to work outside of the home.*

Treatment participants were likewise asked to indicate the number of statements (from 0 to 4) that they agreed with. We recover the true share of participants supporting WWOH by subtracting the average number of stated agreements to the three non-WWOH statements (given by the control group mean) from the average number of stated agreements to the list including the WWOH statement (given by the treatment group mean).

Next, we directly elicited private opinions about WWOH by asking participants in the control group whether they agreed with the statement about the right of women to work outside the home. This enables us to compare, at the sample-level, the *stated* degree of support for WWOH with the degree of support for WWOH *inferred* using the list experiment. Since the latter should be much less susceptible to social desirability bias, this comparison reveals the magnitude of any potential social desirability bias effects.

Finally, we ask participants to gauge the responses of other participants to the statement about WWOH. In the control condition, participants are asked to estimate the percentage of other participants who *reported* agreeing with the same statement. Participants are informed that the study consists of a nationally representative sample of 1,500 married Saudi males aged 18–35. As in the main experiment, this elicitation was conducted in an incentivized manner—participants were informed that the respondent with the most accurate guess would receive a \$50 Amazon gift card.

Since we do not directly elicit private beliefs about the WWOH statement in the treatment condition we cannot elicit participants' beliefs about others in the same way (without revealing the existence of our treatment and control conditions). Instead, we ask participants to estimate the percentage of other participants who would *privately* agree with the WWOH statement.³³ Participants are again informed about the characteristics of the sample. Comparing the distribution of these estimates of the *true opinions* of others to the distribution of estimates of the *answers*

³³This was done without incentives since there is no obvious measure of accuracy available.

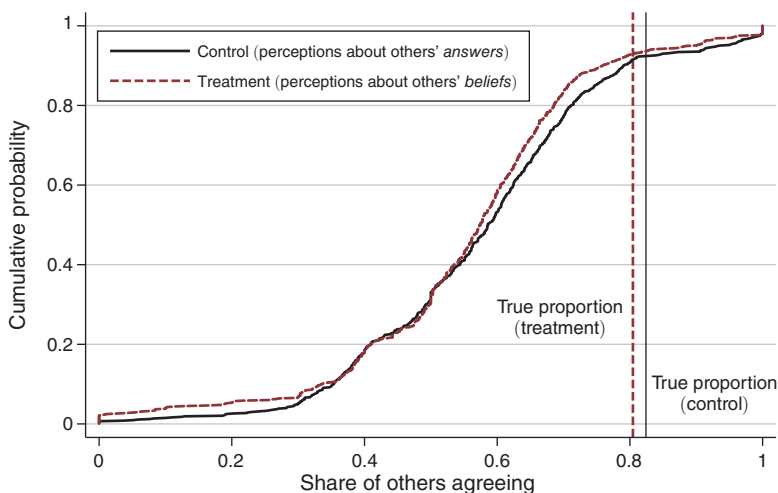


FIGURE 6. MISPERCEPTIONS ABOUT OTHERS' BELIEFS: NATIONAL SURVEY

Notes: CDF of respondents' guesses about the share of other participants in the national survey agreeing with the statement that women should be able to work outside the home. Vertical lines show the true proportion of respondents agreeing with the statement. In the treatment group, private beliefs are elicited using a list experiment. The true proportion of respondents agreeing to the statement in the treatment group is then inferred by subtracting the average number of agreements in the control list from the treatment list (a list of statements identical to the control list with the single addition of the sensitive WWOH statement).

of others allows us to examine whether guesses about the answers of others are distorted by expectations that others might answer strategically.

A pilot sample of 100 participants was launched on February 15, 2018 and concluded on February 16, 2018. No changes were made to the experimental design after reviewing the pilot results. The survey was then administered to the rest of the sample (1,496 survey participants, 1,460 of which completed the whole survey) starting on February 17, 2018 and concluding on March 2, 2018. We drop the data from the pilot in our analysis. Online Appendix Table B9 displays summary statistics for the national survey sample and shows that covariates are balanced across conditions.

Results.—We first assess the external validity of the finding that most Saudi men privately support WWOH while failing to understand that others do. As displayed in Figure 6, in this broader, more-representative sample, 82 percent of men agree with the same statement on WWOH used in the main experiment regarding working outside the home. When incentivized to guess the responses of other survey respondents, 92 percent of them underestimate the true share. This number is higher than in the experimental sample, perhaps because they are no longer being asked about their own neighbors' opinions. That said, the average wedge is similar to before: 25 percentage points.

Next, we use the survey to examine whether social desirability bias/experimenter demand effects could have been a driver of the misperception finding in the main experiment. These results are also shown in Figure 6. By subtracting the average number of agreements in the control list from the treatment list, we get the share

of respondents who agree with the statement of interest, but under response conditions with a higher degree of plausible deniability. We find a very similar share of 80 percent, indicating that social desirability bias/experimenter demand effects are not a driver of the finding. Finally, we examine whether individuals might be incorrectly expecting others to strategically respond to the WWOH agreement question, which would distort guesses about others, since the question asked about how others *answered* the question. We find that beliefs about other participants' *true opinions* were extremely similar to the guesses about others' *answers*.

Finally, we again find that more accurate guesses (for both types of questions regarding others) are correlated with more confidence in the guess, as depicted in online Appendix Figure B8.

Evidence from the Arab Barometer.—As a last check on the external validity of the fact we document, we find that the share of Saudi men supportive of WWOH outside the home is also extremely similar when using the nationally representative sample from the wave of the *Arab Barometer* survey containing that question for Saudi Arabia (2010–2011). Out of approximately 700 male respondents, 77 percent agree with the statement, “A married woman can work outside the home if she wishes.” Among male respondents aged 18–35 (the age bracket in our study), the share is 79 percent.³⁴ The *Arab Barometer* survey also allows us to establish that older men are also supportive of WWOH outside the home; among those over 35, the share agreeing with that statement is 74 percent. Figure 7 displays the share of men supportive of WWOH outside the home in the different samples discussed in this paper.

B. Discussion: Why Are Social Norms Misperceived?

Our results paint a consistent picture. When social norms in the labor market are misperceived, simple information provision can affect labor supply decisions. In the case of Saudi Arabia, most people perceive society to be more conservative than it really is. As a result, too few women work outside of home.

This begs the question, why is society in a misperceptions equilibrium to begin with? It is beyond the scope of this paper to provide sharp answers here, but one may speculate that a lack of communication between individuals in society is a likely underlying mechanism. This mechanism may be particularly relevant if mass media does not appropriately convey to the public what values dominate in society. In any case, our data points to individual-to-individual communication being a relevant underlying mechanism. In the first experiment, the more men in the room the participants know, the more correct perceptions they have (see online Appendix Figure B2). This is consistent with better information arising from communication with people you know. However, this correlation may also simply reflect homophily. That is, it is easier to predict the beliefs of people whom you know because they are more likely to be like you, even in the absence of communication on the issue.

³⁴ The question in the *Arab Barometer* survey is not worded exactly the same as in our survey, and so participants may interpret its meaning slightly differently. Also, participants were asked to report their extent of agreement: strongly agree/agree/disagree/strongly disagree. We pool the first two to create the indicator of agreement.

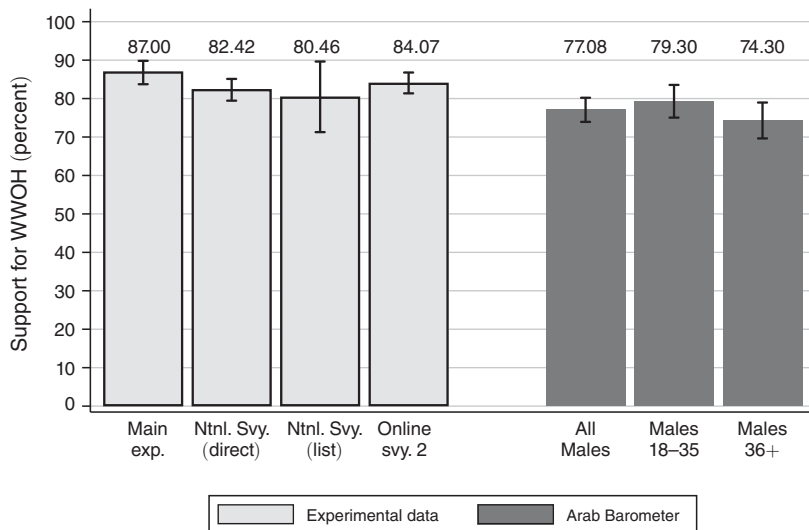


FIGURE 7. SUPPORT FOR WOMEN WORKING OUTSIDE THE HOME ACROSS SAMPLES

Notes: Comparison of support for female labor force participation in experimental data and the nationally representative *Arab Barometer*. In the first three experimental conditions, participants were presented with the following statement, “In my opinion, women should be allowed to work outside the home.” The last experimental condition randomly presents one of two versions of the question. One-half of the sample gets the standard statement as above and the other half were asked “which of the following two statements do you agree with? 1. In my opinion, women should be allowed to work outside of the home. 2. In my opinion, women should not be allowed to work outside of the home if she wishes.” The plot shows the percentage of respondents agreeing to the presented statement in each case. In the *Arab Barometer* survey data, we define the agreement indicator by pooling respondents who reported “agreeing” or “strongly agreeing.” *Ntnl. Svy. (Direct)* refers to the direct elicitation of beliefs conducted in our own nationally representative online survey conducted in early 2018. *Ntnl. Svy. (List)* refers to the level of support inferred from the list experiment in the same survey by subtracting the average number of agreements in the control group from the average number in the treatment group. *Online Svy. 2* refers to the level of support in the second national survey conducted in early 2020 (we pool the answers from the two types of ways the question was asked); 95 percent confidence intervals. Confidence interval for the level of support inferred from the list experiment calculated by taking the variance of the treatment: control estimator to be the sum of the variance of the sample mean in the control group and the variance of the sample mean in the treatment group.

To probe more directly for lack of communication being an important driver of misperceptions, we use data from a separate online survey, implemented in January and February 2020, with a national sample of married Saudi men aged 18–45 (we increased the upper bound on age to make it easier to reach out to a larger sample). A total of 703 individuals were surveyed (online Appendix Table B10 presents sample characteristics). When asked whether women should be allowed to work outside the home, the responses mirror those of the larger national survey: 84 percent are in favor, but the average perception is that 56 percent are in favor; 82 percent of the sample underestimates the level of support. In the age bracket 18–35, the numbers are very similar: 83 percent are in favor, the average perception is 54 percent, and 84 percent underestimate support for WWOH.³⁵

³⁵To rule out that some people mechanically, or mindlessly, answer positively to a normative question regardless of its meaning, in this survey we randomize two versions of how the question was asked. One-half of the sample get the standard question and the other half were asked “Which of the following two statements do you agree with?”

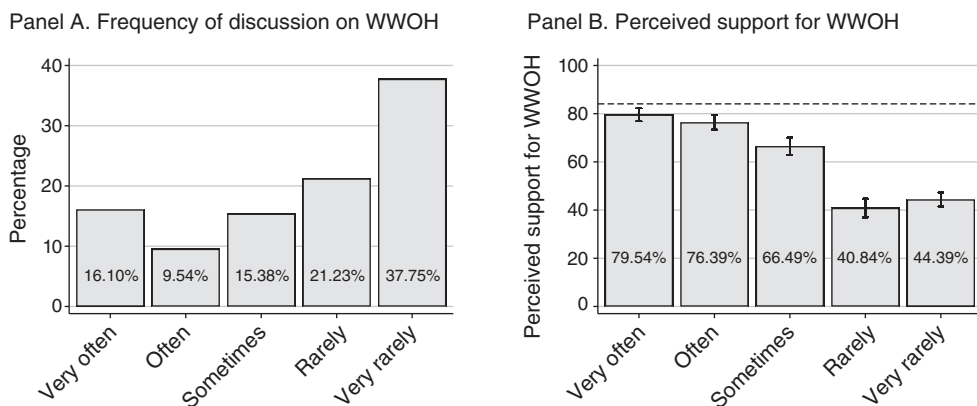


FIGURE 8. FREQUENCY OF DISCUSSION AND PERCEIVED SUPPORT FOR WWOH: SECOND NATIONAL SURVEY

Notes: Panel A shows the distribution of answers to the question, “Think of the topic on whether women should be allowed to work outside of the home. Is this a topic that is discussed [*frequency*] among your male friends and relatives?” Panel B shows the average perceptions of support by others for women working outside the home for each category of answers to this question. The dashed horizontal line shows the true proportion of respondents agreeing with the statement in the sample; 95 percent confidence intervals. The data are from the online survey conducted in early 2020.

It is worth noting that the results in this survey with respect to private opinions regarding WWOH, and beliefs about the opinions of other people, are similar to our other samples. Put differently, as online Appendix Figure B9 makes clear, when we examine the degree of misperceptions across different samples and measurement approaches, a very consistent picture emerges: the vast majority underestimate the degree of support for WWOH and the average wedge is around 25–28 percentage points.

Now, most importantly, in this survey another question is added, “Think of the topic on whether women should be allowed to work outside of the home. Is this a topic that is [very often, often, sometimes, rarely, very rarely] discussed among your male friends and relatives?” First, as seen in panel A of Figure 8, 59 percent of the sample report to discuss the topic either rarely or very rarely (38 percent report to discuss it very rarely). Moreover, if lack of communication helps fuel misperceptions, one would expect a correlation between the frequency of communication and perceptions. Panel B of Figure 8 plots the relationship. Among those that very rarely discuss the issue with their male friends and relatives, the average wedge is 40 percent, while it is 4 percent for those who communicate very often. The relationship between misperceptions and frequency of communication is essentially monotonic across the five frequencies.

This evidence, while suggestive, is thus consistent with a lack of communication driving misperceptions. That said, it is only a simple correlation between actions and beliefs. It is entirely possible that the correlation reflects a form of reverse causality: if I perceive most others to be conservative, so that I (falsely) fear a social stigma from conveying that I am a progressive, then I may refrain from discussing

1. In my opinion, women should be allowed to work outside of the home. 2. In my opinion, women should not be allowed to work outside of the home.” We get very similar levels of support regardless of how the question is asked, with 86 percent support in the first version, and 82 percent support in the second version.

the topic at all. This is a possibility, at least if people dislike lying about their own views to their male friends and relatives. Note that this interpretation is also consistent with the logic of pluralistic ignorance. In this case, it is easy to see how misperceptions may persist for a long time, but it also means that we are back to the original question of why misperceptions exist to begin with. Ultimately, without experimental data that tackles the question directly, it is very difficult to pin down the underlying mechanism driving misperceptions. In our view, it is a fruitful avenue for future research, in the present context and beyond.

IV. Surveyor Recruitment Experiment

A. Design

We partnered with a large Saudi survey company and embedded an experiment in their recruitment of temporary enumerators, conducted in January and February 2020. The company first contacted women in their database (across three cities: Jeddah, Riyadh, and Dammam), following a script we provided them to have harmonized demographic information, and to confirm their interest in an opportunity for a temporary (one-day) enumerator job, calling respondents from home. The pay rate is SAR 25 per hour (or 200 Saudi Riyals per day), the standard one for this type of work.

Our experimental intervention occurred during a second call to women who had confirmed their interest in the first call. In the second call, the company offered to 291 women, aged 18–45, the choice between

- performing the job from home at the previously agreed upon pay rate of SAR 200 per day;
- performing the same enumerator job (for the same survey), but conducting face-to-face interviews in shopping malls. This second option offers a pay that is 20 percent higher, SAR 240 per day, and on top of that, the company will pay for transportation costs.

The second option is more profitable (under reasonable assumptions regarding the opportunity cost of time for commuting to malls), but it involves working outside the home, which is potentially stigmatized. Note that subjects are told that they will interview men in both cases.

Our randomization of interest was at the individual level, right before potential enumerators were offered the choice just described. One-half of the subjects (control group), were read the following passage:

Before I describe the position I would like to tell you that our company supports Vision 2030, which encourages women to participate in the Saudi labor force.

The other half of the subjects (information group), were read the following passage:

Before I describe the position I would like to tell you that our company supports Vision 2030, which encourages women to participate in the Saudi labor force. We would like to share some information about a recent study that may be of interest to you. In a recent survey of a national sample of

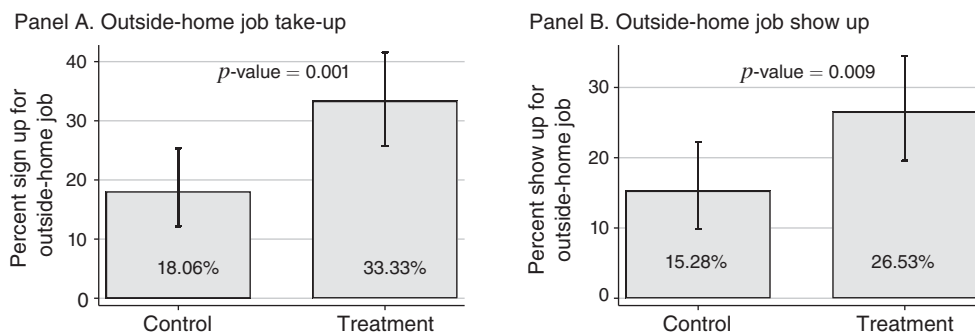


FIGURE 9. SHARE OF WOMEN CHOOSING AND SHOWING UP FOR OUTSIDE-HOME JOB: RECRUITMENT EXPERIMENT

Notes: Labor supply outcomes of female participants in the recruitment experiment; 95 percent binomial proportion confidence intervals. p -values calculated from testing for equality of proportions. Panel A refers to the proportion of women who chose the offer to work outside the home. Panel B refers to the proportion of women who chose the offer and showed up for the outside-the-home job (the outcome is a dummy equal to one if the woman accepted the offer to work outside the home and showed up for the job).

about 1,500 married Saudi men aged 18–35, 82 percent agreed with the statement ‘In my opinion, women should be allowed to work outside of the home.’ This means that the vast majority of young married Saudi men support women working outside of the home.

The two groups get information that makes clear the stance of the company with respect to female labor force participation in Saudi Arabia. However, the second group gets additional information: they learn that the vast majority of young married Saudi men support working outside the home. We therefore use the information generated by our own national survey, conducted previously.

This design let us keep the two positions as similar as possible, while varying exactly the one element we are interested in: whether or not the job is to be performed outside the home. We examine two (pre-registered) outcomes: subjects’ choice (on the phone call) between the two versions of the job and administrative data on job attendance.

B. Results

In online Appendix Table B11 we provide sample characteristics and tests of balance of covariates. Figure 9 displays our main findings. Panel A of Figure 9 shows that in the control condition, 18 percent of participants choose the outside-the-home version of the job. In the group receiving information on the level of support for WWOH by Saudi males, the share increases by 15 percentage points, going to 33 percent (the p -value of a t -test of equality is 0.001). We also create a dummy variable using administrative data from the survey data, equal to one if the woman showed up to perform the outside-the-home job, and zero otherwise. The findings are displayed in panel B of Figure 9. Job attendance behavior provides a consistent picture: 15 percent show up to perform the job outside the home in the absence of the information provision, and 27 percent do so in the group receiving information on social norms (p -value = 0.009). Table 6 displays the findings in regression format.

TABLE 6—SHARE OF WOMEN CHOOSING AND SHOWING UP FOR OUTSIDE-HOME JOB: RECRUITMENT EXPERIMENT

	Outside-home sign up (1)	Outside-home sign up (2)	Outside-home show up (3)	Outside-home show up (4)
Treatment (β)	0.153 (0.0506)	0.157 (0.0514)	0.113 (0.0473)	0.119 (0.0482)
Constant	0.181 (0.0322)	0.147 (0.136)	0.153 (0.0301)	0.180 (0.126)
<i>Inference robustness</i> (β)				
<i>p</i> -value: robust SE	0.003	0.002	0.018	0.014
<i>p</i> -value: permutation test	0.002	0.001	0.027	0.013
Controls		✓		✓
Observations	291	291	291	291
R^2	0.0305	0.0548	0.0191	0.0363

Notes: Columns 1 and 3 report estimates from OLS regressions respectively of an indicator for whether the respondent signed up (column 1) and showed up (column 3) for the job outside the home on the treatment indicator (information that most young married Saudi males support women working outside of home). Columns 2 and 4 include socioeconomic controls for age, marital status, number of children, education, and employment status. Robust standard errors reported in parentheses. Reported *p*-values for permutation tests derived from running 1,000 replications in each case.

Finally, it is important to note that the temporary nature of the job might interact with its potential visibility. It is unclear whether the effects would be larger or smaller if the job was longer, in which case both visibility and monetary payoffs would be greater.

V. Conclusion

In this paper, we provide evidence that (i) the vast majority of married men in Saudi Arabia privately support women working outside of home; (ii) they substantially underestimate the level of support by other men, including their own neighbors; and (iii) correcting these beliefs about others affects their wives' labor supply decisions. These results might help us understand the role that social norms play in constraining job opportunities for women in Saudi Arabia, but also how these constraints might be lifted by the simple provision of information. Active information provision might be particularly important in less democratic societies, where the availability of other natural aggregators of information (such as elections, referenda and even opinion polls) is more limited.

We end with potential avenues for future research. We view our findings as “proof of concept” of the potential for information provision to change behavior regarding female labor supply in Saudi Arabia (and potentially in other countries). We believe that expanding the scale and observing how information spreads in networks, and how it affects a large set of outcomes is an important topic for future work. Also, the goal of the study was to understand the opinions and perceptions of male guardians in Saudi Arabia. As a result, we did not examine opinions and perceptions of women, though our recruitment experiment targeted women. Evidence from the *Arab Barometer* suggests that the vast majority of women in the country are supportive of WWOH (92 percent). Future work can enrich the analysis

on the women's side of the labor supply decision process. Finally, understanding what is at the root of the stigma associated with women working outside of home might help design policies to address it: what are husbands trying to signal to others by acting in opposition to women working outside the home?

REFERENCES

- Acemoglu, Daron, and Matthew O. Jackson. 2017. "Social Norms and the Enforcement of Laws." *Journal of the European Economic Association* 15 (2): 245–95.
- Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53.
- Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics* 128 (2): 469–530.
- Ali, S. Nageeb, and Roland Benabou. 2016. "Image Versus Information: Changing Societal Norms and Optimal Privacy." NBER Working Paper 22203.
- Arab Barometer. 2010–2011. "Arab Barometer Wave II." <https://www.arabbarometer.org/waves/arab-barometer-wave-ii/> (accessed March 2017).
- Baldiga, Katherine. 2014. "Gender Differences in Willingness to Guess." *Management Science* 60 (2): 434–48.
- Bénabou, Roland, and Jean Tirole. 2011. "Laws and Norms." NBER Working Paper 17579.
- Bernhardt, Arielle, Erica Field, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer-Moore. 2018. "Male Social Status and Women's Work." *AEA Papers and Proceedings* 108: 363–67.
- Bertrand, Marianne. 2011. "New Perspectives on Gender." In *Handbook of Labor Economics*, Vol. 4B, edited by David Card and Orley Ashenfelter, 1543–90. San Diego: Elsevier.
- Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130 (2): 571–614.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. "Beliefs about Gender." *American Economic Review* 109 (3): 739–73.
- Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. Forthcoming. "From Extreme to Mainstream: The Erosion of Social Norms." *American Economic Review*.
- Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107 (11): 3288–3319.
- Bursztyn, Leonardo, Alessandra González and David Yanagizawa-Drott. 2017. "Understanding and Shifting Social Norms of Female Labor Force Participation in Saudi Arabia." AEA RCT Registry. October 02. <https://www.socialscienceregistry.org/trials/2447>.
- Bursztyn, Leonardo, Alessandra González and David Yanagizawa-Drott. 2018. "Understanding and Shifting Social Norms of Female Labor Force Participation in Saudi Arabia: Follow-Up." AEA RCT Registry. January 08. <https://www.socialscienceregistry.org/trials/2633>.
- Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020a. "Replication Data for: Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." American Economic Association. <https://doi.org/10.3886/E119404V1>.
- Bursztyn, Leonardo, Alessandra González and David Yanagizawa-Drott. 2020b. "Understanding and Shifting Social Norms of Female Labor Force Participation in Saudi Arabia: Additional Experiment." AEA RCT Registry. January 18. <https://www.socialscienceregistry.org/trials/5321>.
- Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *Quarterly Journal of Economics* 130 (3): 1329–67.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–27.
- Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." *Quarterly Journal of Economics* 129 (4): 1625–60.
- Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the Conventional Thought that Counts: How Third-order Inference Produces Status Advantage." *American Sociological Review* 82 (2): 297–327.
- DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56.
- DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2017. "Voting to Tell Others." *Review of Economic Studies* 84 (1): 143–81.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–50.

- Eckel, Catherine C., and Philip J. Grossman. 2008. "Men, Women and Risk Aversion: Experimental Evidence." In *Handbook of Experimental Economics Results*, Volume 1, edited by Charles R. Plott and Vernon L. Smith, 1061–73. San Diego: Elsevier.
- Enikolopov, Ruben, Alexey Makarin, Maria Petrova, and Leonid Polishchuk. 2020. "Social Image, Networks, and Protest Participation." Unpublished.
- Fernández, Raquel. 2007. "Alfred Marshall Lecture: Women, Work, and Culture." *Journal of the European Economic Association* 5 (2–3): 305–32.
- Fernández, Raquel. 2013. "Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century." *American Economic Review* 103 (1): 472–500.
- Fernández, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1 (1): 146–77.
- Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. "Mothers and Sons: Preference Formation and Female Labor Force Dynamics." *Quarterly Journal of Economics* 119 (4): 1249–99.
- Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore. 2016. "On Her Account: Can Strengthening Women's Financial Control Boost Female Labor Supply?" Unpublished.
- General Authority for Statistics, Kingdom of Saudi Arabia. 2017. "Labor Force Survey." <https://www.stats.gov.sa/en> (accessed March 2017).
- Giuliano, Paola. 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association* 5 (5): 927–52.
- Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4): 1091–1119.
- González, Alessandra L. 2013. *Islamic Feminism in Kuwait: The Politics and Paradoxes*. New York: Palgrave Macmillan.
- International Labour Organization. 2017. "ILOSTAT Database." <https://ilostat.ilo.org/data/> (accessed March 2017).
- Jayachandran, Seema. 2015. "The Roots of Gender Inequality in Developing Countries." In *Annual Review of Economics*, Volume 7, edited by Kenneth J. Arrow and Timothy F. Bresnahan, 63–88. Palo Alto, CA: Annual Reviews.
- Jensen, Robert, and Emily Oster. 2009. "The Power of TV: Cable Television and Women's Status in India." *Quarterly Journal of Economics* 124 (3): 1057–94.
- Katz, Daniel, Floyd H. Allport, and Margaret B. Jenness. 1931. *Students' Attitudes: A Report of the Syracuse University Reaction Study*. New York: Craftsman Press.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Kuran, Timur. 1991. "The East European Revolution of 1989: Is It Surprising That We Were Surprised?" *American Economic Review* 81 (2): 121–25.
- Kuran, Timur. 1995. *Private Truths, Public Lies*. Cambridge: Harvard University Press.
- Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.
- List, John A., Azeem M. Shaikh, and Yang Xu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22 (4): 773–93.
- McKelway, Madeline. 2019. "Vicious and Virtuous Cycles: Self-Efficacy and Employment of Women in India." Unpublished.
- O'Gorman, Hubert J. 1975. "Pluralistic Ignorance and White Estimates of White Support for Racial Segregation." *Public Opinion Quarterly* 39 (3): 313–30.
- Perez-Truglia, Ricardo, and Guillermo Cruces. 2017. "Partisan Interactions: Evidence from a Field Experiment in the United States." *Journal of Political Economy* 125 (4): 1208–43.
- Prentice, Deborah A., and Dale T. Miller. 1993. "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm." *Journal of Personality and Social Psychology* 64 (2): 243–56.
- Raghavarao, Damaraju, and Walter T. Federer. 1979. "Block Total Response as an Alternative to the Randomized Response Method in Surveys." *Journal of the Royal Statistical Society: Series B (Methodological)* 41 (1): 40–45.
- Tankard, Margaret E., and Elizabeth Levy Paluck. 2016. "Norm Perception as a Vehicle for Social Change." *Social Issues and Policy Review* 10 (1): 181–211.
- Voigtländer, Nico, and Hans-Joachim Voth. 2015. "Nazi Indoctrination and Anti-Semitic Beliefs in Germany." *Proceedings of the National Academy of Sciences* 112 (26): 7931–36.